Computer typesetting as input to information systems

J. H. Kuney

An important element of system design is knowing where a process starts and where it ends. This is not as simple an analysis as it first seems. For the moment we can assume that the process starts with a manuscript to be processed into a printed page. It is already clear that as new devices now being developed become available for use, we will have to do some rethinking about the starting point of the typesetting process.

The output stage of a hot metal typesetting system is type which serves but one purpose, the printing of an impression on a sheet of paper or other surface, the type to be stored for later reprinting. It can be corrected or rearranged for a use other than that of the original setting, but this is more an exception than the ordinary use of type in printing.

When computers are used in the typesetting process, typesetting becomes a by-product of the process of coding the manuscript. Instead of the keystroke leading to the selection of a matrix for casting in hot metal or on film, the keystroke now results in holes on a paper tape, or as an alignment of electrons on magnetic tape, either of which the computer can read and manipulate at very high speeds. This very basic change in the result of the keyboarding process opens up new approaches to the role of the computer in publishing.

At a recent discussion sponsored by the National Science Foundation on *Computer Applications in Scientific Publication*, the chairman stated in his introductory remarks:

'The basic concept of a single keyboarding of machine-recording of text material at the manuscript stage is fraught with exciting implications. Consider, if you will, the machine-recording of text material at the manuscript stage of all necessary elements of data and information that will permit, in addition to the composition of the article itself, the subsequent derivation of secondary information—such as indexes, abstracts, bibliographic listings, and the availability in machine-usable form of such information for storage and search on mechanized retrieval systems.'

Computerized typesetting is but one element of a total scientific information system, a system which is just beginning to take form and where the coded tape produced in the keyboarding of manuscripts may be used in all subsequent phases of disseminating and handling scientific information.

Our goal is to develop computer programs which can handle the variety of formats and special characters required in the composition of scientific manuscripts, including text, tables, mathematical formulae, and chemical structures. In our present computer

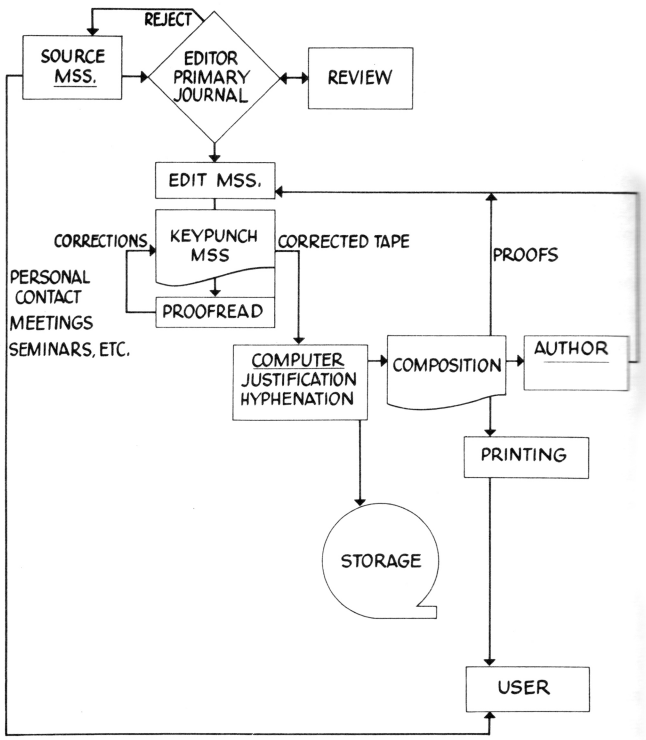

Fig 1 Flow of manuscript handling for primary publication using computer-aided typesetting parallels other composition methods

typesetting system, this ability is limited only by the capability of our photocomposing equipment, a Model 200 Photon equipped with an 8-channel tape reader. The Photon matrix disc provides 1,440 characters, each available in 12 different sizes, line lengths up to seven inches, and almost unlimited variations of line spacing in increments of a tenth of a point.

Already faster and more versatile typesetting equipment is being produced. With this thought in mind, we have kept our three main system components—input, processing, and output—as independent as possible. Each can thus be developed separately as new equipment is introduced and programming capability is improved. In addition, we decided that handling procedures for manuscripts should parallel those used in hot-metal processing in order to minimize the introduction of new procedures or of added problems related to man-machine interaction. The editorial review of the manuscript and its preparation for the compositor is handled in the same manner as for other composition methods. In the first step of the computer-aided typesetting process, the manuscript is keyboarded on a Friden Flexowriter (Programmatic) without justification or hyphenation instructions to produce both a hard copy and an 8-channel punched

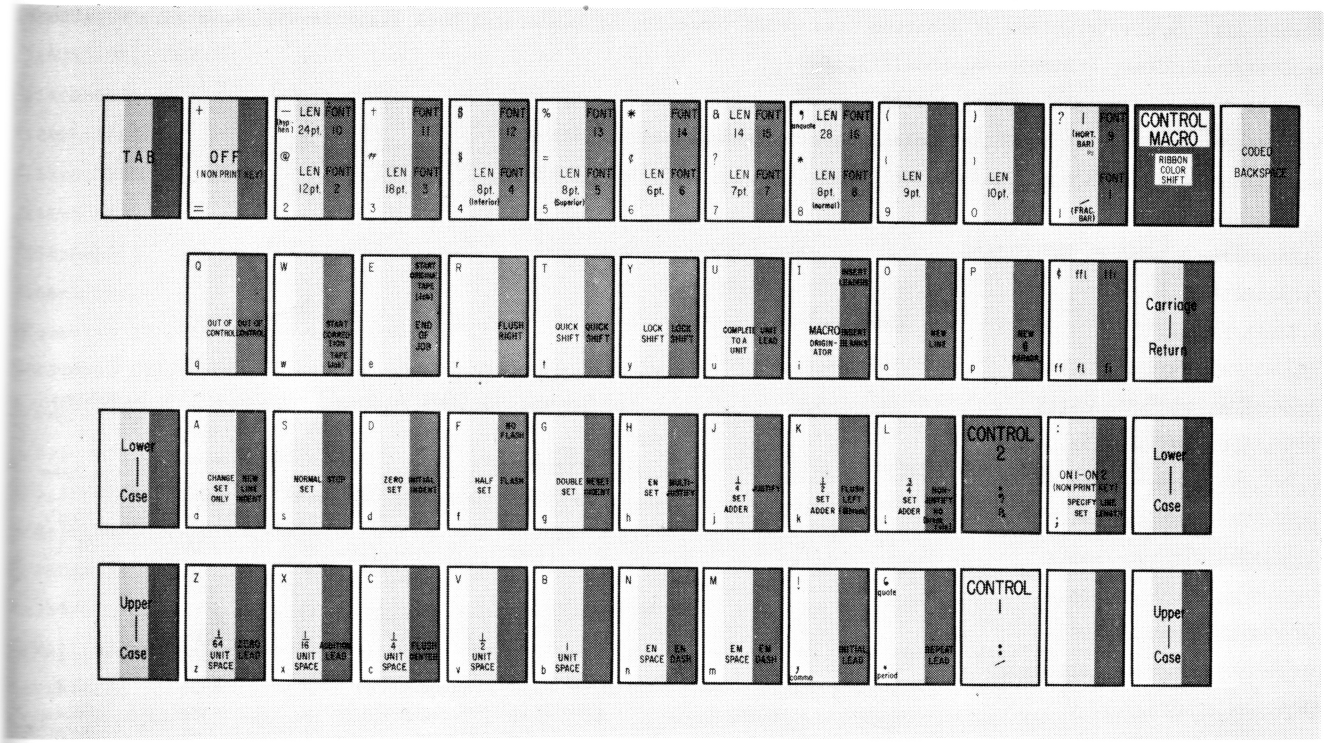

Fig 2 Friden keyboard layout showing use of control keys

paper tape in Flexowriter code (a binary-coded decimal system with odd parity). As part of the manuscript encoding, the operator includes typesetting instructions to the computer to control type style, type size, line spacing, line length, special characters, and all of the other variables which are involved in the format of a scientific journal.

The layout of the keyboard is based on the standard 45-key typewriter format of a Friden Flexowriter. With the use of the upper case shift the character capacity is doubled to the ninety characters in a font on the matrix disc of the Photon. To provide a unique key for each of the three computer controls and still average a single keystroke for each character to be typeset, the ligatures were moved from a character position to a control position on the keyboard.

Two controls are used to signal the computer that the characters that follow are instructions to the computer programs. Control 1 for the most part covers lens changes, set, and spacing functions. Control 2 is for font changes, leading, line length, and end-of-line signals. Each of these functions is obtained by punching the proper control key, followed by a single character with the exception of leading and line length.

The majority of the functions are found in the lower case positions. Where the upper case position is shown as a blank, the function is available in either the upper case or the lower case position. To minimize the use of the case shift when selecting functions in the opposite case mode, a 'quick shift' function is used to obtain a single character in a new function and case shift, followed by automatic return to the original operating mode. A function is also available to make a permanent shift to a new function and remain in this condition until the operator brings it back manually. Using these half-turn shifts, the number of keystrokes is decreased, the operator's procedures are simplified, and all manual case shifts in the input stage could be eliminated. However, we retain the case shifts for text characters in order to have them appear in the hard copy.

Control 3 is obtained by punching the ribbon colour shift, followed by the macro-identification and a second ribbon colour shift. Control 3 is used to express macro-function which, when read in the computer, signals into operation one or more subroutines stored in the computer programs and/or recalls stored textual material. Control 3 is obtained by punching the ribbon colour shift, followed by the macro-identification and a return to the black ribbon position. The alphabetic characters 'a' to 'z' are reserved for permanent macros. These are pre-punched and stored in the computer under a journal identification and are recalled for use from storage when the identification of a special journal is used.

In the ACS system of computerized typesetting, input coding has been simplified through the extensive use of macro-codes which enable the operator to use fewer keystrokes when entering typesetting instructions or repetitive text matter. For example, in the original keyboarding of the title for an article in a journal, the operator used the ribbon colour shift key to put the keyboard in the macro-control condition. Next he hit the lower case 'a', the assigned macro-code for a title setting. After a return to the black-ribbon position, the copy for the title was punched. When processed by the computer program, the output tape carried the typesetting instructions to set a two-line centred heading in 18 pt Century Schoolbook, with proper space between the lines, and 30 pt of film advance after setting the second line. The instructions also included rules for proper breaking of the two lines.

```
hTable III. /x/xDistribution of Notation Size (4383 Compounds of Commercial Deck Assigned Wiswesser Line Notations)
&x050&3/6&c&,030&091&hSize&c&061&h&c&122Size
&c&091&,080&hr&c&061&h&c&122&hr&c&044&3
&c&044&hDesigna&c&047&h&c&061&hCumulative&c&hDesigna&c&h&cCumulative
&c&044&htion (;#s&3/x)&c&047&hNo.&c&061&h=&c&htion (;#S&3/x)&c&hNo.&c=
/11&c&044&h
/12&c&047&h   } MACRO INSTRUCTIONS
/13&c&061&h
&,100&3/8&u&u
①1②/n1③/n0.1③D③14&c/n89.6
①2②/n2③/n1.8③E③15&c/n91.3
①3②/n3③10.4③F③16&c/n92.9&x050
①4②/n4③19.2③G③17&c/n94.1
①5②/n5③29.5③H③18&c/n94.7
①6②/n6③39.6③I③19&c/n95.4
①7②/n7③56.8③J③20&c/n95.9
①8②/n8③60.3③K③21&c/n96.4
①9②/n9③67.8③L③22&c/n96.6&x050
①0②10③74.5③M③2&c/n97.0
①A②11③80.0③N;NZ③24&n36&c100.0
①B②12③84.3③&c
①C②13③87.5③&c
&e
```

Fig 3 Hard copy showing input for setting of table. Temporary macro-codes (circled) control positioning of table elements

Table III. Distribution of Notation Size (4383 Compounds of Commercial Deck Assigned Wiswesser Line Notations)

Size			Size		
Designation (S)	No.	Cumulative %	Designation (S)	No.	Cumulative %
1	1	0.1	D	14	89.6
2	2	1.8	E	15	91.3
3	3	10.4	F	16	92.9
4	4	19.2	G	17	94.1
5	5	29.5	H	18	94.7
6	6	39.6	I	19	95.4
7	7	56.8	J	20	95.9
8	8	60.3	K	21	96.4
9	9	67.8	L	22	96.6
0	10	74.5	M	23	97.0
A	11	80.0	N–Z	24–36	100.0
B	12	84.3			
C	13	87.5			

Fig 4 Computer-set table produced from input shown in Fig 3

The permanent macro-codes apply to the repetitive format patterns which characterize the typographic style of a particular journal. In addition, it is possible for the keyboard operator to enter temporary macros to meet the special needs of a particular article or series of articles. For example, if a long chemical term appeared frequently in a particular manuscript, the operator could enter the term as a temporary macro. Thereafter a single keystroke, plus the two control shifts, would provide the complete term in the output tape. In effect, the operator uses the macro-codes as a typesetting language to reduce input effort and to optimize computer processing.

The punched paper tape is read into the IBM 1460, using an IBM 1012 paper tape reader/punch. The computer processes control modes and case shifts and the result is written out on a magnetic tape. During this process, a master file of job tapes is built from which all further processing is initiated.

From the master file, a computer print-out with each line numbered is prepared. After proofreading of the print-out, corrections are set on a new punched tape and merged into the original job in the master file. The updated master file is produced by batch processing in which both corrections and new job tapes may be intermixed. Working on the corrected job tape, the processor now adds justification, hyphenation and format decisions.

The final output component of the system consists of generating a punched tape to control the photocomposing machine. The control tape is an 8-channel paper tape containing two frames of codes for each character. The first frame carries the character code and the second frame provides the spacing information. The control tape is read directly into the photo unit of the photocomposing machine and the end product after processing of the exposed film is a positive film in galley form. Proofs are sent to the author and upon return corrections are keyboarded on a correction tape and merged with the original on the master file. Either the corrected tape is completely rerun through the main processor and through the Photon, or only the corrected lines can be processed and then stripped into the original film. Makeup is handled manually by stripping the film on a transparent base and then sent to the printer for manufacture into a plate for offset printing.

The pattern of decision-making, which goes into the setting of straight text matter, is varied but relatively straightforward when compared to that required in setting scientific manuscripts. Here spacing considerations and character selection become much more complex, and for processing via computer it is necessary to reduce these spatial relationships to some linear form for input coding.

The scope of these problems can best be illustrated by an actual example of the keyboarding process for the setting of a mathematical equation using a computer. In the setting, four lens settings and seven fonts will be used as follows: lens—normal 8 pt, slanted numerator 8 pt, slanted denominator 8pt, and 18 pt

```
&Etest,j000,d000,gl,p,bl&o&o&o
/1g&j&3/9&,100&346
/11&z&r&170
/12&z&c&355
/13&z&c&385
/14&z&c&422
/15&z&r&256
/16&z&r&346
(3)Equation 17 will reduce to&x150&o
(1);¢d;#I&3  ;¢e;=n &1/3$/fT/4;=/d0/3&1;F/fT&f/s/x/5&5L/3&1'
(2)/8w(2)/5&5N(2)/4;=j  0
(3)/8;¢u;=j
(4)/5;$2/d9/s;#f;=j(4)/8&3/f&m/f&m/f&m/s/r(4)/4;$2;#y;=n
(5)/8;¢o;=n/3&1&m/8;#d1/3&1=
(6)/8&3(21)
&x150
(3)/mAt the optimal value of &t3I, when the extremal is not restricted, the value ;¢d;#I&3 must vanish for all
variations ;¢o;=n&3(&t3I).It is therefore seen from Equation 21 that a necessary condition for an optimum is&o
&e
```

MACRO INSTRUCTIONS

Fig 5 Hard copy of input for setting of mathematics. Temporary macros (circled) control positioning of formula elements

Equation 17 will reduce to

$$\delta I' = \epsilon_n \left[\int_0^L \left(\sum_j^N v_j \frac{\partial \bar{f}_j}{\partial y_n} \omega_n \right) dl \right] \qquad (21)$$

At the optimal value of I', when the extremal is not restricted, the value $\delta I'$ must vanish for all variations $\omega_n(l)$ It is therefore seen from Equation 21 that a necessary condition for an optimum is

Fig 6 Computer-set formula produced from input shown in Fig 5

for the two-line characters; font changes—italic, Greek, Roman, maths, caps and small caps, superior and inferior.

After the operator has determined that the equation will fit in the assigned column width, the equation is divided into segments based on spacing treatment. Thus the segments are divided as follows: flush right, centre, justify, centre, and flush left. The sixth segment is not included since the key number sets flush right on the over-all column width. Note the need to set up segment 2 separately because of the need to centre the limits over the sigma.

After the operator has determined the line lengths of each of the segments, either by calculation or by estimation, he is now ready to begin punching the input tape. First temporary macros defining the spacing requirements of the respective segments are entered. In the actual setting, the macro is followed by the keying of the actual characters to be used in the segment. After processing in the computer, the output tape is run through the Photon. Since preparation of this paper, the process has been further simplified through the use of tab codes: the operator is able to enter the copy in a more conventional manner but still leave the spacing calculations to the computer.

The use of a computer in the production of typeset chemical structures is of particular importance in our work, both as a typesetting problem and as a vital element in the handling of chemical information. In an earlier experiment, we set structures on the Photon using a direct keystroke-to-character approach; that is, for each keystroke on the input device, we generated a character on the Photon. This was successful but somewhat unsophisticated. Our present approach is to 'describe' the structure to the computer and let the computer process the 'description' into output for the typesetting of the structure.

As input we use the paper tapes produced by chemical structure typewriters which produce codes to represent normal alphanumeric symbols, special symbols to represent chemical bonds, and coordinates which supply a frame of reference within which to locate both text and graphic data. No decision-making is required on the part of the chemical structure typist other than to reproduce accurately the structure provided as copy. Typically, coordinates are reproduced automatically by the typewriter at the end of a string of input symbols. The end of a string is defined by changing the spatial reference of the input by rolling the platen, moving the carriage more than one space right, backspacing, and returning the carriage. The coordinates refer to the location of the beginning of the preceding symbol string. When processed by the computer, the output is a paper tape to set the chemical structure in line-by-line order. The setting is accomplished on the Photon, using a special disc containing the necessary characters to set structures. The disc is now in the manufacture stage and the computer programs are being written.

We have now assembled the elements necessary to handle the composition of scientific journals. But in terms of system design,

typesetting is but one element. We have only begun to attack such problems as the proper handling of manuscripts for computer processing, proofreading, authors' corrections, page formatting, storage of work in progress, and storage of tapes of completed journals, all of which must be solved to take best advantage of computer processing capability. We also must be alert to the changing technology and its possible impact on the design of our systems.

It is a long step from our present output rate of somewhat under eight characters per second to the 500 and more per second promised by photocomposing equipment using character generation techniques. At least one such device is already in use and several others are nearing the prototype stage. In this equipment characters are generated from the memory of the computer rather than by passing light through a character matrix. Characters are built up of dots, squares, or line segments on the face of a cathode ray tube and then exposed on to sensitized paper as direct computer printout or film. Estimates of speed of 500 characters per second and up for graphic arts quality and speeds as high as 8,000 characters per second for quality suitable for proofreading purposes are predicted.

In addition, the release from the limitation of a fixed matrix grid suggests practically unlimited availability of type styles and special characters.

Optical character readers are increasing in use where the number of different text characters required is limited, for example, the editing and proofreading of manuscripts utilizing light pen techniques; that is, cathode ray tube systems which permit information to be written directly on the face of the tube.

It is very interesting to explore the feasibility and practicability of including, at the time of the first processing of manuscripts, the data and information required if the original record is to find use in the manufacture of secondary publications such as abstract journals, bibliographies and indexes, as well as in the storage, search, and retrieval of scientific information.

To do this the computer must recognize those parts of the original input which have been selected as having potential value in secondary processing. This would include material which might be added at the input stage such as index terms. In the ACS system the previously described macro-code used for typesetting instructions serves more importantly as a flag to identify respective portions of the manuscript for computer use in search and retrieval. Thus the macro-code which issues typesetting instructions for the format of the title of journal articles also identifies or flags that part of the manuscript as a title within the coded record.

Identification as a title is unique even when the same macro-code is used in producing a variety of formats among different journals. By searching on the macro-code used for title setting, a computer could generate a listing of titles. Similar logic would apply in the case of authors, abstracts, literature citations; in fact, the whole gamut of identifiable manuscript levels which might be flagged for content and/or for typographic format.

A national system

In the US a national system of handling chemical information has been proposed as a necessary requirement in meeting the nation's needs for research progress. In such a system, publishers and editors of chemical journals would continue to exercise control over the content of the material they admit into the information system. Once accepted for publication, manuscripts are sent to the information-processing centre. At this point the manuscripts are analysed for their potential role in the total information system and the necessary storage, search and retrieval data are added. The manuscript is then encoded. The publisher may receive a tape for processing through his own equipment for publication, or the centre may provide publication service through the point of preparing the finished film ready for platemaking. Still another possibility is that the centre will act as a transmittal centre either in the form of printed journals or by some form of facsimile transmission whereby the scientist would but press a button in his laboratory or office to receive the material.

Such thinking is no longer visionary. The equipment is available and many of the components of the total system are being tested. But there remains the significant task of putting them together and testing the acceptability and economic value of systems such as the one proposed here. If such systems do have a place, what does it mean for the printers of scientific journals? I cannot give an answer today other than to suggest that now is the time to study and to seek understanding of the implications of the changes which are taking place.

Acknowledgement

We wish to express our appreciation of the National Science Foundation for their support of this work under NSF Grants GN 140 and GN 426.

Literature

Proceedings Computer Applications in Scientific Publication Washington DC. February 1966.

Kuney, J.H., Lazorchak, B.G. Machine set chemical structures. *Proceedings of the American Documentation Institute*, **1**, 1964.

Cossum, W.E., Hardenbrook, M.E. Computer generation of atom-bond connection tables from hand-drawn chemical structures. *Proceedings of The American Documentation Institute* **1**, 1964.